Genome-wide association study of microscopic colitis in the UK Biobank confirms immune-related pathogenesis

Harry D Green [1,2], Robin N Beaumont [1], Amanda Thomas [2], Benjamin Hamilton [2], Andrew R Wood [1], Seth Sharp [1], Samuel E Jones [1], Jessica Tyrrell [1], Gareth Walker [2], James Goodhand [2], Nicholas A Kennedy [2], Tariq Ahmad [2], Michael N Weedon [1]*

1 - Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK
2 - IBD Pharmacogenetics, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Short Title: Microscopic Colitis Genome-Wide Association Study

* Corresponding Author: Michael N. Weedon

M.N.Weedon@exeter.ac.uk
University of Exeter Medical School
Medical Research - Level 3, RILD Building
Royal Devon & Exeter NHS Foundation Trust (Wonford)
Barrack Road
Exeter, EX2 5DW
United Kingdom

Abstract

Background and Aims: The causes of microscopic colitis are currently poorly understood. Previous reports have found clinical associations with coeliac disease and genetic associations at the HLA locus on the ancestral 8.1 haplotype. We investigated pharmacological and genetic factors associated with microscopic colitis in the UK Biobank.

Methods: 483 European UK Biobank participants were identified by ICD10 coding, and a genome-wide association study was performed using BOLT-LMM, with a sensitivity analysis performed excluding potential confounders. The HLA*IMP:02 algorithm was used to estimate allele frequency at 11 classical human leukocyte antigen (HLA) genes, and downstream analysis was performed using FUMA. Genetic overlap with inflammatory bowel disease (Crohn's disease and ulcerative colitis) was investigated using genetic risk scores.

Results: We found significant phenotypic associations with smoking status, coeliac disease and the use of proton-pump inhibitors but not with other commonly reported pharmacological risk factors. Using the largest sample size to date, we confirmed a recently reported association with the MHC Ancestral 8.1 Haplotype. Downstream analysis suggests association with digestive tract morphogenesis. By calculating genetic risk scores, we also report suggestive evidence of shared genetic risk with Crohn's disease, but not with ulcerative colitis.

Conclusions: This report confirms the role of genetic determinants in the HLA in the pathogenesis of microscopic colitis. The genetic overlap with Crohn's disease suggests a common underlying mechanism of disease.

1. Introduction

Microscopic colitis includes two related inflammatory bowel disorders, lymphocytic colitis and collagenous colitis that have a combined prevalence of 103 cases per 100,000 population [1]. Both disorders cause chronic watery non-bloody diarrhoea and incontinence, and are associated with normal endoscopic appearances and characteristic histological features. The primary histological feature of lymphocytic colitis is patchy lymphocytic infiltration of the epithelium with preserved crypt architecture. Collagenous colitis is characterised by the presence of a thickened subepithelial collagen layer.

The pathogenesis of microscopic colitis is poorly elucidated: it reportedly involves immune responses to luminal factors in genetically predisposed individuals [2]. A recent association study based on Immunochip data reported association between HLA alleles on the 8.1 haplotype and collagenous colitis [3] but not with lymphocytic colitis [4] in cohorts comprising 314 patients with collagenous colitis, 122 patients with lymphocytic colitis and 4,299 controls. Further, Westerlind et. al. reports genetic overlap with inflammatory bowel disease (IBD) by comparing the number of nominally significant Single Nucleotide Polymorphisms (SNPs) in both phenotypes [3]. The most frequently cited environmental risk factors are medications, with non-steroidal anti-inflammatory drugs (NSAIDs), proton-pump inhibitors (PPIs) and selective serotonin reuptake inhibitors (SSRIs) most commonly implicated [5–7]. There have been no reported investigations of potential pharmacogenetic risk factors for microscopic colitis.

We sought phenotypic and genetic associations with microscopic colitis in individuals of European ancestry enrolled in the UK Biobank, and report subsequent downstream analysis. Subsequently, we stratified the data by drug use and performed a GWAS to identify pharmacogenetic associations. Finally, we calculated genetic risk scores for IBD (Inflammatory Bowel Disease) to quantify in detail genetic overlap with microscopic colitis.

2. Methods

2.1 Participants

The UK Biobank is a population-based prospective study comprising more than 500,000 UK participants aged 40-69 years at time of recruitment between 2006 and 2010. Participants are actively followed, and phenotype data collected includes demographics, medical conditions, medications, lifestyle and anthropometric measurements. SNP genotypes were generated from the Affymetrix Axiom UK Biobank array (~450,000 individuals) and the UK BiLEVE array (~50,000 individuals). More detail on the UK Biobank can be found elsewhere [8].

We defined microscopic colitis by the ICD10 code K52.8. The full UK Biobank dataset contains 522 K52.8 cases and 502,097 controls (104.0/100,000 persons). We stratified the data to contain only white Europeans by performing principal component analysis in the 1000 Genomes reference panel. Of the 451,099 participants remaining, 483 are cases and 450,616 are controls (107.2/100,000 persons). Of the 483 cases, 335 (69.4%) had K52.8 as their primary diagnosis code.

We investigated pharmacological associations with NSAIDs, PPIs and SSRIs using the medication codes reported in the UK Biobank (Data-Field 20003: Treatment/medication code). Drugs included and their relative frequencies in the UK Biobank among cases and controls can be found in supplementary table 1.

2.2 Statistical Methods

We performed tests for association with clinical characteristics using the Mann Whitney U test for continuous data and Fisher's exact test for categorical data. P values are reported uncorrected for multiple testing, but we used an adjusted p-value threshold of 0.0045 for statistical significance for the clinical data.

Quality control of the genotype data was performed centrally by the UK Biobank.[9] For the genome-wide association study, we used ~12.0M Haplotype Reference Consortium (HRC) imputed variants with an imputation $r^2 \geq 0.9$, minor allele frequency (MAF) ≥ 0.025 (2.5%) and with a Hardy–Weinberg equilibrium $p > 1 \times 10^{-12}$. Further, individuals with IBD or coeliac disease were excluded from all genetic analyses, leaving 423 cases and 445,232 controls.

We performed our main association test using BOLT-LMM v2.3 [10], which applies a linear mixed model (LMM) to adjust for the effects of population structure and individual relatedness and allowed us to include all related individuals in our white European subset, rather than reducing the sample size to only include the unrelated individuals (379,768). Covariates included were age, sex, recruiting centre and genotyping chip. A more detailed explanation of GWAS methodology can be found in our recent publication [11], and a principal components plot is available in Supplementary Figure 1. Odds Ratios from BOLT-LMM were calculated by $OR = e^{\beta / (\mu * (1 - \mu))}$ where μ = case fraction, and standard errors divided by $(\mu * (1 - \mu))$ to give confidence intervals. Following the GWAS, FUMA's SNP2GENE analysis was used to convert SNP data into genomic loci, and perform a gene-set enrichment analysis using MAGMA [12].

As a sensitivity analysis we also performed a secondary GWAS on a more refined phenotype using only white British unrelated individuals, in which we excluded coeliac and inflammatory bowel disease (IBD) participants (defined by ICD10, ICD9 and Self Report) and participants using PPIs . Controls were further refined by excluding those with a diagnosis (self-reported or in HES data) before the age of 70 of coronary heart disease, stroke, diabetes, chronic obstructive pulmonary disease, renal failure, any cancer (excluding non-melanoma skin cancer), and those who had died from any cause before the age of 70. To further limit the possibility of type 1 errors, this GWAS was performed using Fisher's exact test. This analysis used 335 cases and 64,300 controls (253 cases and 49,608 controls following filtering related individuals).

Imputation of HLA alleles was performed using the HLA*IMP:02 algorithm to estimate allele frequency at 11 classical human leukocyte antigen (HLA) genes: HLA-A, -B, -C, -DRB5, -DRB4, -DRB3, -DRB1, -DQB1, -DQA1, -DPB1 and -DPA1 using reference panels described by Motyer et al [13]. This procedure provides accurate dosages for 362 SNPs in the HLA region, allowing identification of genetic basis of autoimmune processes with greater precision[14].

Genetic risk scores for ulcerative Colitis (UC), Crohn's disease (CD) and IBD were calculated using odds ratios for previously published SNPs [15]. From that study, using only SNPs that were genome-wide significant at 5×10^{-8} and also present in our imputation panel, we have 145 SNPs for CD, 89 for UC and 162 for IBD. To quantify genetic overlap, for each IBD phenotype, we compared the mean GRS among the microscopic colitis patients with controls (defined as not having UC, CD, IBD or microscopic colitis). A GRS containing N SNPs was calculated according to the equation below where β is the β-coefficient (log odds ratio) representing the association between each SNP and the relevant phenotype and d_i is the estimated dosage. Statistical associations are quantified using a 2-tailed t-test.

$$GRS = \sum_{i=1}^{N} \beta_i \cdot d_i$$

3. Results

3.1 Clinical Associations

Variable	Cases	Controls	*p*-value
Age at recruitment (years)	**61.9 [56.2 – 65.4]**	**58.6 [50.5 – 63.8]**	**5×10^{-15}**
Body mass index	26.6 [23.7 – 29.6]	26.7 [24.1 – 29.9]	0.195
Townsend Deprivation Index	-2.03 [-3.55 – 0.28]	-2.27 [-3.70 – 0.23]	0.171
Female	**65.6% [317/483]**	**54.2% [244,531/450,616]**	**5×10^{-7}**
Current smoker	**14.7% [71/483]**	**10.4% [46,792/450,616]**	**0.003**
Coeliac disease	**3.3% [16/483]**	**0.4% [1,991/450,616]**	**7×10^{-10}**
IBD	**9.5% [46/483]**	**0.8% [3,896/450,616]**	**2×10^{-32}**
Using PPIs	**20.3% [98/483]**	**10.3% [46,397/450,616]**	**9×10^{-11}**
Using SSRIs	5.8% [28/483]	4.0% [18,003/450,616]	0.048
Using NSAIDs	28.4% [137/483]	26.3% [118,687/450,616]	0.326
Using Statins	19.7% [388/483]	16.3% [73,365/450,616]	0.048

*Table 1: **Clinical Associations with Microscopic Colitis** – Table showing demographic and drug use associations in the UK Biobank. Continuous variables are reported in terms of median [interquartile range]. All p values were computed by a Fisher's or Mann-Whitney U test. p value threshold was 0.0045 after Bonferroni correction. Significant associations are in bold.*

Demographic and drug associations are shown in Table 1. Microscopic colitis patients were more likely to be older, female and to be a current smoker. We find white European UK Biobank participants with microscopic colitis had an 8 times higher risk of coeliac disease and 12 times higher risk of IBD than controls. We found no evidence of association with body mass index (BMI), or socioeconomic status (defined by Townsend Deprivation Index).

In terms of drug associations, patients with microscopic colitis were 2-times more likely to be using PPIs (a sub-analysis suggests 1.95 for omeprazole, 2 times for lansoprazole). We found no significant associations for SSRIS, NSAIDs or statins when accounting for multiple testing. We also performed a sub-analysis of NSAIDs, stratifying by COX1 and COX2 inhibitors, but found no significant associations. All clinical associations were robust to adjustment for age and sex in a multivariable logistic regression model, and testing with or without the related individuals.

3.2 GWAS Results

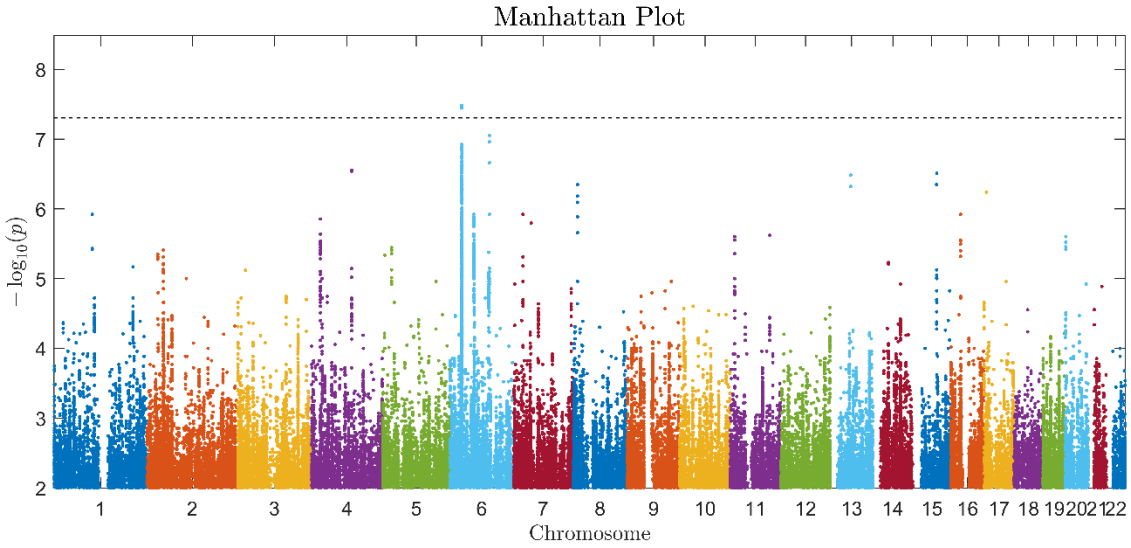


*Figure 1: **Manhattan Plot** – A plot showing the $-\log_{10}(p)$ values for each SNP in the HRC Imputation Panel with p<0.01, computed using Bolt-LMM and plotted using an in-house MATLAB script which we have made publicly available on the MATLAB File Exchange [16]. The horizontal dashed line is the genome wide significance threshold at $p=5\times10^{-8}$. The strongly significant signal on Chromosome 6 lies in the HLA region, with lead SNP rs2596560, $p=3\times10^{-8}$.*

Figure 1 shows a Manhattan plot of all SNPs that passed QC and had p<0.01. We find a strong association for microscopic colitis on chromosome 6, in the HLA region, with lead SNP rs2596560 (OR 0.64, 95% confidence interval 0.56-0.72, $p=3\times10^{-8}$, MAF 0.24 vs 0.33). The odds ratio remained the same at the lead SNP (OR 0.64, 95% confidence interval 0.53-0.77, $p=3.3\times10^{-6}$) when we performed our secondary analysis, on the refined phenotype using Fisher's exact test

Summary statistics for all genomic risk loci with $p<10^{-5}$ are available in Supplementary Table 2, and the full GWAS summary statistics can be found at https://www.ebi.ac.uk/gwas/home. In Supplementary Figure 2 we present a QQ plot. The genomics inflation factor λ was 1.0000. We also include the top 10 results of a MAGMA Gene-set Analysis in Supplementary Table 3, in which we find a Bonferroni-significant association for digestive tract morphogenesis.

SNP	Allele Frequency	Beta	Standard Error	*p*-value
B_801	**0.143**	**4.3×10^{-4}**	**9.4×10^{-5}**	**3.9×10^{-6}**
DRB1_301	**0.147**	**3.7×10^{-4}**	**9.2×10^{-5}**	**5.6×10^{-5}**
DRB3_9901	0.655	-2.6×10^{-4}	6.9×10^{-5}	1.5×10^{-4}
DRB3_101	**0.166**	**3.3×10^{-4}**	**8.8×10^{-5}**	**2.0×10^{-4}**
DQA1_501	**0.230**	**2.9×10^{-4}**	**7.8×10^{-5}**	**2.1×10^{-4}**
C_701	**0.175**	**3.2×10^{-4}**	**8.6×10^{-5}**	**2.4×10^{-4}**
DQB1_201	**0.148**	**3.4×10^{-4}**	**9.2×10^{-5}**	**2.7×10^{-4}**

*Table 2: **Classical HLA Alleles** - A table showing the allele frequencies, beta, standard error and p values of all classical HLA alleles with a p value below 5×10^{-4} (computed through Bolt-LMM) and MAF >0.01. SNPs in bold are part of the HLA A1-B8-DR3-DQ2 (ancestral MHC 8.1) haplotype.*

HLA imputation demonstrated that the SNP rs2596560 associates with the class I and II alleles that comprise the ancestral MHC 8.1 haplotype previously been linked with microscopic colitis, with lead

SNP B_0801 passing genome-wide significance. Results from BOLT -LMM runs on the HLA imputed alleles are given in Table 2, with alleles on the MHC 8.1 haplotype highlighted in bold.

To follow up on the phenotypic association with PPIs, we sought pharmacogenetic associations by using PPIs as an inclusion criterion (comparing PPIs with microscopic colitis and PPIs with no microscopic colitis). However, there were no genome-wide significant associations for this GWAS.

3.3 Genetic Overlap with IBD

Table 3 shows the mean Crohn's disease, ulcerative colitis and IBD genetic risk scores and 95% confidence intervals for microscopic colitis patients and controls.

Disease for which genetic risk score derived	**Mean (95% CI) score for microscopic colitis patients**	**Mean (95% CI) score for controls**	**p value**
Crohn's Disease	0.9688 (0.9632-0.9739)	0.9634 (0.9632-0.9636)	0.035
Ulcerative Colitis	0.9928 (0.9859-0.9997)	0.9890 (0.9888-0.9892)	0.261
IBD	0.9286 (0.9237-0.9335)	0.9230 (0.9229-0.9231)	0.019

*Table 3: **Genetic Risk Score Means and Associations** - Microscopic colitis patients are significantly enriched for genetic risk factors for Crohn's disease and IBD, but not for ulcerative colitis*

Microscopic colitis patients had a higher genetic risk for all three tests, but only Crohn's disease (p=0.035) and IBD (p=0.019) were significant at the 5% confidence level, suggesting some shared genetic pathway behind microscopic colitis and Crohn's disease / IBD. These were robust to applying the same tests on only the unrelated individuals in the UK biobank as a sensitivity analysis. In Table 4 we show which of the known risk loci replicate at p<0.05 for microscopic colitis, although none of these are significant when Bonferonni correcting the p value threshold.

GRS	**Chr**	**Pos**	**A1**	**A1 Freq**	**Beta**	**SE**	**P**
Both	10	6081230	C	0.836092	2.2×10^{-4}	8.85×10^{-5}	1.4×10^{-2}
Both	2	28614794	C	0.519615	1.5×10^{-4}	6.58×10^{-5}	1.9×10^{-2}
Both	22	30493882	G	0.55509	1.8×10^{-4}	6.58×10^{-5}	7.7×10^{-3}
Both	17	32593665	A	0.72546	1.6×10^{-4}	7.32×10^{-5}	3.0×10^{-2}
Both	21	45615741	G	0.394392	1.3×10^{-4}	6.69×10^{-5}	4.7×10^{-2}
Both	7	50175654	G	0.573233	-1.7×10^{-4}	6.62×10^{-5}	1.2×10^{-2}
Both	7	50304461	C	0.33364	-1.4×10^{-4}	6.94×10^{-5}	4.1×10^{-2}
Crohn's Disease	12	6491125	G	0.646309	-1.6×10^{-4}	6.93×10^{-5}	1.8×10^{-2}
IBD	7	2869985	T	0.697662	2.3×10^{-4}	7.12×10^{-5}	1.1×10^{-3}
IBD	13	27531267	T	0.82362	2.2×10^{-4}	8.58×10^{-5}	1.1×10^{-2}

*Table 4: **Nominally Significant SNPs in Crohn's Disease and IBD GRS** – Summary statistics of the known risk variants for Crohn's disease and IBD in our microscopic colitis GWAS for those that pass the nominal p value threshold of 0.05.*

4. Discussion

We have conducted a clinical and genetic case-control study of microscopic colitis in the UK Biobank. We confirm previously reported phenotypic associations with age, sex, coeliac disease, smoking status and PPIs. In a genome-wide association study, we have confirmed recent reports of association with SNPs on the MHC 8.1 haplotype, indicating an immune component to the

pathogenesis of microscopic colitis. Using genetic risk scores, we obtain results consistent with overlap in genetic risk factors for Crohn's disease and IBD but not ulcerative colitis. This may suggest shared genetic pathways between these phenotypes.

This is the largest genome-wide association study of microscopic colitis to date, and the first to look at genetic overlap between microscopic colitis and IBD by using a genetic risk score. The main limitation of the study is that the ICD10 coding in the UK Biobank only covers the first decimal place: K52.8. We acknowledge two limitations as a consequence: this code also includes eosinophilic gastritis and colitis, and that we are unable to distinguish lymphocytic from collagenous colitis. However, eosinophilic gastritis and colitis are of very low prevalence (5.1 and 2.1 /100,000 persons respectively)[18] compared to microscopic colitis (103.0/100,000 persons)[1].

The UK biobank ICD10 data relies on hospital coding data, which has limitations in terms of whether patients have had a coded diagnosis at hospital and relies on the accuracy of hospital coding; however, their use is standard practice in literature performing UK Biobank GWASs. This study was only performed using white Europeans, and further studies are required to determine if results are consistent across other ethnic groups. The UK Biobank also only includes patients between the ages of 40 and 69 at recruitment, although this covers the peak age of onset for microscopic colitis. We are confident our main GWAS result is not a false positive due to the high MAF of the lead SNP and aligning closely with previous work, but we acknowledge that due to having under 500 cases, there may by many SNPs we were unable to detect due to low odds ratio or MAF. A meta-analysis combining the results of this study and others may help identify further associations.

To follow up on the strong association with PPIs, we performed a GWAS to find possible underlying pharmacogenetic associations, but there were no genome-wide significant SNPs. A follow-up study with a dedicated drug-exposed cohort would have greater power to detect such associations, as has been demonstrated in our recent study of thiopurine-induced myelosuppression[19].

References

1. Storr MA. Microscopic Colitis: Epidemiology, Pathophysiology, Diagnosis and Current Management—An Update 2013. *ISRN Gastroenterol*. 2013;2013:1-12. doi:10.1155/2013/352718

2. Ianiro G. Microscopic colitis. *World J Gastroenterol*. 2012;18(43):6206. doi:10.3748/wjg.v18.i43.6206

3. Westerlind H, Mellander M-R, Bresso F, et al. Dense genotyping of immune-related loci identifies HLA variants associated with increased risk of collagenous colitis. *Gut*. 2017;66(3):421-428. doi:10.1136/gutjnl-2015-309934

4. Westerlind H, Bonfiglio F, Mellander M-R, et al. HLA Associations Distinguish Collagenous From Lymphocytic Colitis. *Am J Gastroenterol*. 2016;111(8):1211-1213. doi:10.1038/ajg.2016.215

5. Riddell RH, Tanaka M, Mazzoleni G. Non-steroidal anti-inflammatory drugs as a possible cause of collagenous colitis: a case-control study. *Gut*. 1992;33(5):683-686. http://www.ncbi.nlm.nih.gov/pubmed/1612488.

6. Masclee GMC, Coloma PM, Kuipers EJ, Sturkenboom MCJM. Increased risk of microscopic colitis with use of proton pump inhibitors and non-steroidal anti-inflammatory drugs. *Am J Gastroenterol*. 2015;110(5):749-759. doi:10.1038/ajg.2015.119

7. Pascua MF, Kedia P, Weiner MG, Holmes J, Ellenberg J, Lewis JD. Microscopic Colitis and Medication Use. *Clin Med Gastroenterol*. 2010;3:CGast.S4469. doi:10.4137/CGast.S4469

8. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779

9. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*. January 2017. http://biorxiv.org/content/early/2017/07/20/166298.abstract.

10. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47(3):284-290. doi:10.1038/ng.3190

11. Jones SE, Lane JM, Wood AR, et al. Genome-wide association analyses of chronotype in 697,828 individuals provides new insights into circadian rhythms in humans and links to disease. *bioRxiv*. January 2018. http://biorxiv.org/content/early/2018/04/19/303941.abstract.

12. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5

13. Motyer A, Vukcevic D, Dilthey A, Donnelly P, McVean G, Leslie S. Practical Use of Methods for Imputation of HLA Alleles from SNP Genotype Data. *bioRxiv*. January 2016. http://biorxiv.org/content/early/2016/12/09/091009.abstract.

14. Dilthey AT, Moutsianas L, Leslie S, McVean G. HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics*. 2011;27(7):968-972. doi:10.1093/bioinformatics/btr061

15. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47(9):979-986. doi:10.1038/ng.3359

16. Green H. Manhattan Plots for visualisation of GWAS results. 2018. https://uk.mathworks.com/matlabcentral/fileexchange/69549-manhattan-plots-for-visualisation-of-gwas-results.

17. Goessling W, North TE, Loewer S, et al. Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. *Cell*. 2009;136(6):1136-1147. doi:10.1016/j.cell.2009.01.015

18. Mansoor E, Saleh MA, Cooper GS. Prevalence of Eosinophilic Gastroenteritis and Colitis in a Population-Based Study, From 2012 to 2017. *Clin Gastroenterol Hepatol*. 2017;15(11):1733-1741. doi:10.1016/j.cgh.2017.05.050

19. Walker GJ, Harrison JW, Heap GA, et al. OC-046 Nudt15 variants contribute to thiopurine-induced myelosuppression in european populations. *Gut*. 2017;66(Suppl 2):A23 LP-A24. http://gut.bmj.com/content/66/Suppl_2/A23.2.abstract.

20. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335-1341. doi:10.1038/s41588-018-0184-y

Funding

S.E.J. is funded by the Medical Research Council (grant: MR/M005070/1) M.N.W. is supported by the Wellcome Trust Institutional Strategic Support Award (WT097835MF). A.R.W. is supported by the European Research Council grants: SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC and 323195. R.B. is funded by the Wellcome Trust and Royal Society grant: 104150/Z/14/Z. J.T. is funded by a Diabetes Research and Wellness Foundation Fellowship. The funders had no influence on study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

None declared.

Author Contributions

H.G., J.G., N.K., T.A., M.W. participated in the conception, design and coordination of the study. H.G., R.B. and M.W. performed data analysis. H.G. and M.W. participated in writing the paper. All authors assisted in the writing, reviewing and approval of the manuscript.

Acknowledgements

This research has been conducted using the UK Biobank Resource (application 9072). The authors would like to acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work.

Supplementary Table 1

Class	Drug	Biobank Code	Cases	Controls
PPIs	Omeprazole	1140865634	11.2% [54/483]	6.4% [28,868/450,616]
	Lansoprazole	1140864752	7.7% [37/483]	3.8% [17,065/450,616]
	Esomeprazole	1141177526	1% [5/483]	0.35% [1,590/450,616]
	Rabeprazole	1141168584	0.4% [2/483]	0.2% [842/450,616]
	Pantoprazole	1140929012	1% [5/483]	0.2% [854/450,616]
SSRIs	Sertraline	1140867878	0.6% [3/483]	0.5% [2,121/450,616]
	Paroxetine	1140867888	0.8% [4/483]	0.3% [1,539/450,616]
	Fluvoxamine	1140879544	0% [0/483]	0.01% [29/450,616]
	Fluoxetine	1140879540	1.9% [9/483]	1.2% [5,436/450,616]
	Escitalopram	1141180212	0.4% [2/483]	0.2% [825/450,616]
	Citalopram	1140921600	2.1% [10/483]	1.8% [8,187/450,616]
NSAIDs	Aceclofenac	1140925806	0% [0/483]	0.01% [54/450,616]
	Aspirin	1140868226	18.4 [89/483]	13.7 [61,826/450,616]
	Celecoxib	1141176662	0.2% [1/483]	0.1% [492/450,616]
	Dexketoprofen	1141164746	0% [0/483]	0% [7/450,616]
	Diclofenac Sodium	1140878036	0% [0/483]	0.07% [307/450,616]
	Etodolac	1140871188	0% [0/483]	0.07% [335/450,616]
	Etoricoxib	1141180140	0% [0/483]	0.09 [407/450,616]
	Fenoprofen	1140871226	0% [0/483]	0% [3/450,616]
	Ibuprofen	1140871310	11% [53/483]	13% [58,536/450,616]
	Ketoprofen	1140871506	0% [0/483]	0.03% [135/450,616]
	Mefenamic Acid	1140871542	0% [0/483]	0.1% [476/450,616]
	Meloxicam	1140926732	0.6% [3/483]	0.3% [1,310/450,616]
	Nabumetone	1140875336	0% [0/483]	0.05% [206/450,616]
	Naproxen	1140871462	0.6% [3/483]	0.74% [3,353/450,616]
	Piroxicam	1140871666	0.2% [1/483]	0.04% [158/450,616]
	Sulindac	1140871604	0% [0/483]	0% [15/450,616]
	Tenoxicam	1140875346	0% [0/483]	0% [13/450,616]
	Tiaprofenic Acid	1140871614	0% [0/483]	0% [8/450,616]
Statins	atorvastatin	1141146234	3.1% [15/483]	3.4% [15,439/450,616]
	fluvastatin	1140888594	0.2% [1/483]	0.05% [213/450,616]
	pravastatin	1140888648	0.4% [2/483]	0.6% [2457/450,616]
	rosuvastatin	1141192410	0.8% [4/483]	0.7% [29,44/450,616]
	simvastatin	1140861958	15.3% [74/483]	11.9% [53,451/450,616]

Supplementary Table 1: ***UK Biobank Medication Codes –*** *The treatment codes in UK Biobank used to define drug phenotypes.*

Supplementary Table 2

CHR	POS	A1	A1FREQ	BETA	SE	P
1	102300829	A	0.78544	-0.00039	7.98E-05	1.20E-06
1	211852006	C	0.895625	0.000483	0.000107	6.80E-06
2	29002900	A	0.96978	-0.00088	0.000192	4.50E-06
2	44211546	A	0.420328	-0.00031	6.65E-05	3.90E-06
3	21580556	T	0.972868	-0.0009	0.000201	7.60E-06
4	24338324	C	0.774897	-0.00038	7.85E-05	1.40E-06
4	108478191	T	0.831665	0.000448	8.72E-05	2.80E-07
5	6222490	C	0.973156	-0.00093	0.000203	4.60E-06
5	24256226	C	0.943636	-0.00066	0.000143	3.60E-06
6	*31355318*	*T*	*0.758228*	*-0.00042*	*7.64E-05*	*3.30E-08*
6	65350792	G	0.83685	0.000415	8.86E-05	2.80E-06
6	104003704	C	0.929096	-0.00057	0.000128	8.20E-06
6	106000790	C	0.951844	-0.00082	0.000153	8.90E-08
7	25573204	G	0.789508	-0.00039	8.04E-05	1.20E-06
7	47391321	G	0.970032	-0.00092	0.000191	1.60E-06
8	12517525	G	0.969694	-0.00097	0.000192	4.50E-07
11	11931183	C	0.950888	-0.00071	0.000152	2.50E-06
11	105858586	A	0.911153	-0.00055	0.000117	2.40E-06
13	54703907	GCA	0.974328	-0.00107	0.00021	3.30E-07
14	41804188	A	0.520016	-0.0003	6.55E-05	5.90E-06
15	64768700	T	0.972247	-0.00102	0.000199	3.10E-07
16	26117039	C	0.871776	-0.00048	9.80E-05	1.20E-06
17	6571439	C	0.964387	-0.00088	0.000176	5.80E-07
20	2374455	G	0.869848	-0.00046	9.80E-05	2.50E-06

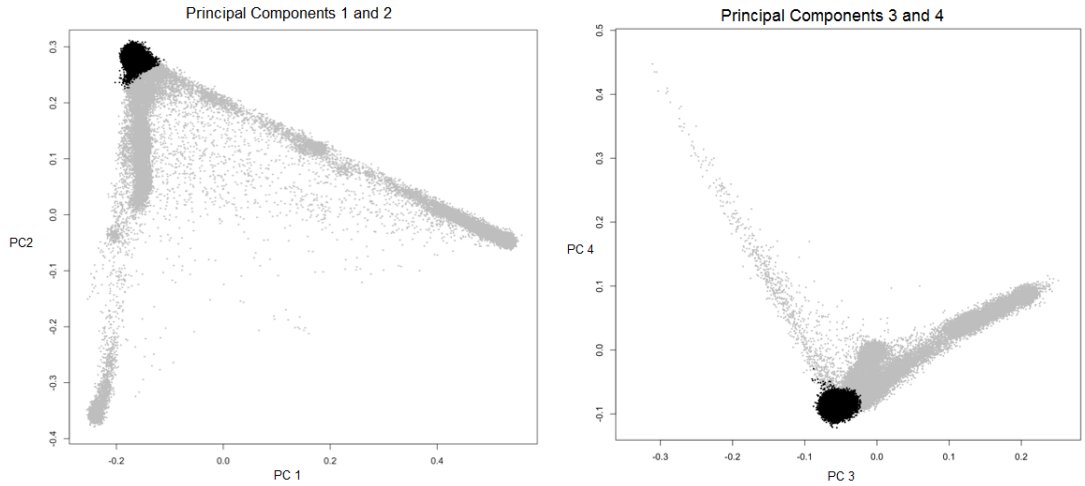
Supplementary Table 2: ***GWAS Summary Statistics –*** *All risk loci with $p<10^{-5}$ from the main GWAS, using all European Biobank Participants and BOLT-LMM. N=483 cases and 450,616 controls. SNPs have been condensed into genomic loci using FUMA* [12]*. Full summary statistics can be found at the GWAS catalog at https://www.ebi.ac.uk/gwas/home.*

Supplementary Table 3

Gene Set	N genes	Beta	Beta STD	SE	P	P_{bon}
GO bp:go digestive tract morphogenesis	48	0.601	0.0303	0.133	3.02E-06	0.032
GO bp:go morphogenesis of an epithelial fold	15	0.802	0.0226	0.212	7.50E-05	0.799
GO bp:go negative regulation of bmp signaling pathway	41	0.493	0.023	0.135	0.000127	1
GO bp:go regulation of calcium ion transmembrane transport	116	0.279	0.0218	0.0769	0.000146	1
GO bp:go prostate gland development	42	0.471	0.0222	0.135	0.000243	1
GO bp:go wnt signaling pathway calcium modulating pathway	39	0.464	0.0211	0.133	0.000245	1
GO mf:go r smad binding	23	0.62	0.0216	0.179	0.000274	1
GO bp:go regulation of calcium ion transmembrane transporter activity	71	0.33	0.0202	0.0957	0.000283	1
GO bp:go pattern specification process	416	0.148	0.0217	0.0431	0.000304	1

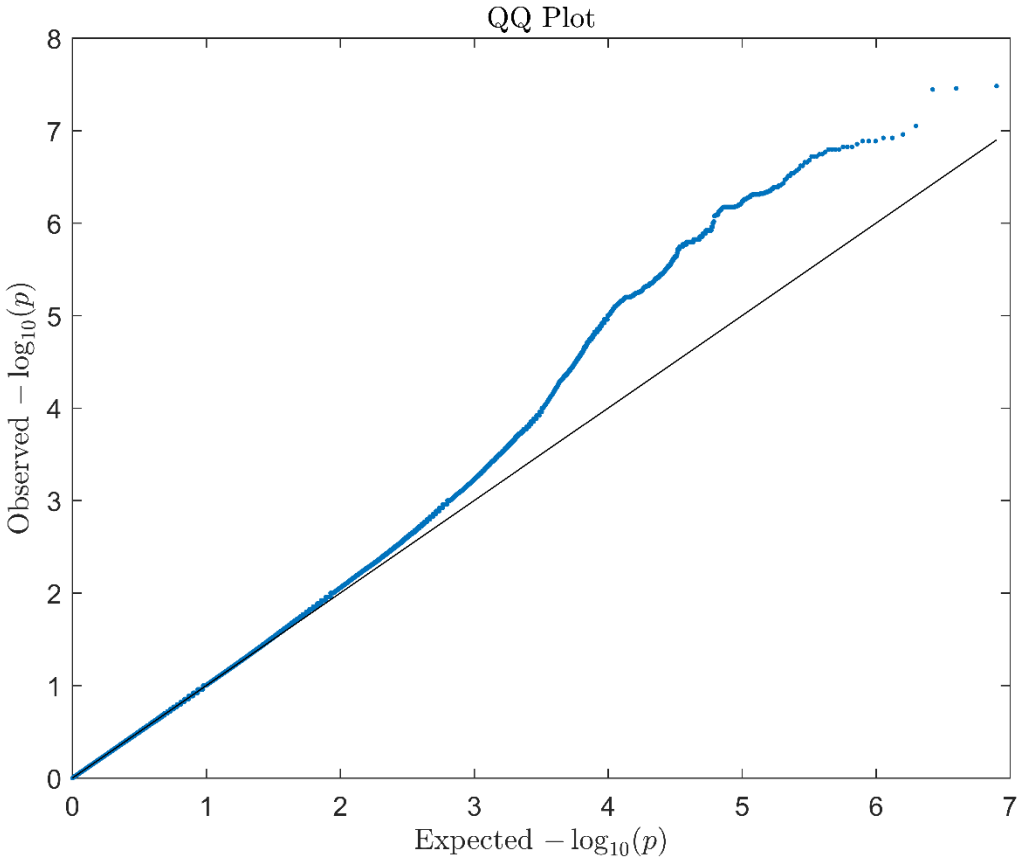
Supplementary Table 3: ***MAGMA Gene-Set Analysis –*** *Top 10 results of a gene set enrichment analysis performed FUMA. We find a significant association for digestive tract morphogenesis when accounting for multiple testing.*

Supplementary Figure 1



Derived European set of individuals after K-means clustering, anchoring on centres of PCs 1-4 for individuals self-reporting as from European population

Supplementary Figure 2



A QQ plot of our main GWAS result using BOLT-LMM on a Biobank Cohort with IBD and coeliac patients excluded. The median p value was 0.5000 and inflation factor λ=1.0000, showing no statistical inflation of p values.